PATENT APPLICATION BASED ON: Docket Number 81360DMW

Inventor(s):   Nathan D. Cahill

Attorney: David M. Woods

Document ID: \DOCKETS\81360

POPULATION MIXTURE MODELING WITH AN INDETERMINATE NUMBER OF SUB-POPULATIONS

04/11/01

POPULATION MIXTURE MODELING WITH AN INDETERMINATE NUMBER OF SUB-POPULATIONS

BACKGROUND OF THE INVENTION

5

Field of the Invention:

The present invention relates generally to data processing. More particularly, the present invention relates to the field of population mixture modeling and threshold segmentation of data.

10

Description of the Related Art:

Optimization problems arise in nearly every area of science and engineering. For example, a chemist may want to determine the minimum energy structure of a certain polymer, a biologist may need to investigate optimal dosages and scheduling of drugs administered to a cancer patient, a civil engineer may need to decide how many elevators to place in an office building so that waiting times are minimized, or an image scientist may wish to be able to automatically locate people in an image. A typical optimization problem takes the form of: minimize *f(x)*; subject to; $x \in D$. Where $x = (x_1, x_2, \ldots, x_n)^T$ is a vector of parameters, and *f* is an objective function. Stated another way: find x* such that $f(x^*) \le f(x)$ for all $x \in D$. Note that the converse problem, maximization problem, can be posed in this manner by simply negating *f(x)*. If *D* is a discrete set, the problem is referred to as a combinatorial optimization problem. On the other hand, if *D* is closed and *f(x)* is continuous over its domain, this is a continuous optimization problem.

A variety of methods exist to determine local solutions for both combinatorial and continuous optimization problems. That is, x' is a local solution if $f(x') \le f(x)$ for all $x \in N \cap D$, where *N* is a neighborhood of x'. For combinatorial problems, a simple choice is the iterative improvement method, which is also properly described as a 'deterministic descent' method. More sophisticated methods, such as a branch and bound method, which is a method that solves a series of

continuous optimization problems, are also known in the art. For continuous

problems, a variety of local minimization methods also exist. Such methods require

function values (simplex), function values and derivatives (conjugate gradient and

quasi-Newton methods), or function values first and second derivatives (Newton and

5   restricted step methods). In situations where D is compact, it can be represented by

the intersection of equality and inequality constraints. Such constrained optimization

problems arise frequently in practice and can be solved in a variety of ways; many of

which are based on the theory of Lagrange multipliers.

The problem of global optimization is much more complicated.

10   Theoretically, global minimization is as simple as finding the minimum of all local

minima and boundary points. In application, this can be impractical. Most local

optimization techniques require an initially feasible point $x^{(1)} \in D$ from which to

begin an algorithmic process of optimization. In order to find a given local minima,

it is necessary to choose a feasible point that will converge to that minima. However,

15   because there are often times many local minima, determining such a feasible point

for a particular local minima can be as difficult as solving the optimization problem

itself.

Global optimization problems can be categorized as either stochastic or

deterministic problems. Stochastic, or Monte Carlo, methods (e.g., random search,

20   Metropolis algorithm, simulated annealing, genetic algorithms, etc.) randomly sample

D in an attempt to locate feasible points that are local minima (or lie sufficiently close

to local minima). Deterministic methods carry out some non-random search

procedure to sift through local minima and find the global minimum. For example,

branching techniques set up simple sub-problems, which are individually solved.

25   Sub-energy tunneling techniques iteratively transform the objective function to

remove local minima. Path following techniques indirectly locate local minima by

solving a system of differential equations.

In the field of digital imaging, optimization techniques are useful as

part of a point-based image segmentation process. Segmenting grayscale images is a

30   useful task in imaging science. Image segmentation can be used directly for image

compression or as a preprocessing step for background removal, feature detection, or pattern recognition algorithms, or possibly other applications. A wide variety of segmentation techniques have emerged in the literature over the last two decades. The simplest methods, and the most efficacious to representation as a global

5   optimization problem, are point-based, where each pixel is quantized based on thresholds determined by analysis of the histogram of the image. There also exist more sophisticated (and time consuming) approaches that are the region-based methods, which take into account local spatial characteristics of the image while quantizing each pixel. Naturally, these techniques are equally applicable to color

10   images.

Generally, point-based image segmentation is comprised a several discrete operations. First a histogram of a digital image is generated. Second, parameters for a plurality of probability density functions are determined, which correspond to a like number of sub-populations in the histogram data. Third,

15   threshold levels are specified segmenting the image data population according to the probability density functions. And fourth, the threshold levels are applied to the histogram. Optimization, or more particularly minimization, occurs with respect to the second operation.

A population mixture model is a probability density function that can

20   be expressed as the weighted sum of multiple sub-populations (other probability density functions). In applications such as image segmentation or cluster analysis, algorithms exist in the prior art that require fitting a population mixture model to a given set of data (a histogram in the case of image segmentation). Current methods for determining the optimal parameters of the population mixture model require prior

25   knowledge of the number of sub-populations appropriate for the data being analyzed. In the prior art, this knowledge is frequently attained by manual inspection or by an automatic pre-processing step. Alternatively, optimization techniques utilizing a multi-start approach have been employed. Such methods fit the model repeatedly for different numbers of sub-populations. Multi-start approaches have two major

30   drawbacks, however. The maximum number of sub-populations is arbitrarily chosen

by some preprocessing step, and the computational effort required to repeatedly fit the model is immense.

Therefore, there is a need in the art to eliminate manual, preprocessing, and other multi-start approaches to population mixture modeling that require *a priori* knowledge of the number of sub-populations appropriate to a given data set prior to conducting an optimization process, or require immense computational effort, like the multistart approach.

SUMMARY OF THE INVENTION

The need in the art is addressed by the systems and methods of the present invention. In accordance with the invention, a method of fitting a plurality of sub-population functions to data is taught. The method comprises the steps of defining a plurality of functions according to a plurality of function parameters and a total number of functions and generating an objective function based on the plurality of function parameters. Then, determining a fitting error between the objective function and the data and comparing the fitting error to stopping criteria, and determining if the fitting error does not satisfy the stopping criteria. Finally, altering the plurality of function parameters and the total number of functions, and repeating the generating, determining, and comparing steps.

In a refinement of the foregoing method, a further step is added which is specifying at least a first threshold value delineating the plurality of functions. In another refinement, the at least a first threshold value is calculated based upon the likelihood of misclassification of data. In another refinement, the step of segmenting the data according to the at least a first threshold value is added. In another refinement, the objective function is defined as a vector representation of the plurality of function parameters. In another refinement, the altering step is accomplished by evolving the plurality of function parameters and the total number of functions according to a genetic algorithm. In another refinement, the genetic algorithm evolves the plurality of function parameters through mutation and crossover. In another refinement, the plurality of functions are normal distributions, and the

plurality of functions parameters include the mean and standard deviations of the normal distributions. In another refinement, the comparing step includes the utilization of a statistical f-test to evaluate the relative contribution of each of the plurality of functions in comparison of the fitting error and the data. In another

5   refinement, the data is organized as a histogram. In another refinement, the stopping criteria are defined by a fitness function. In another refinement, the fitness function is optimized to minimize the magnitude of the fit error between the objective function and the data.

The present invention also teaches an apparatus for fitting a plurality

10   of sub-population functions to data. That apparatus comprises a means for defining a plurality of functions according to a plurality of function parameters and a total number of functions and a means for generating an objective function based on the plurality of function parameters. In addition, a means for determining a fitting error between the objective function and the data and a means for comparing the fitting

15   error to stopping criteria, and wherein if the comparison determines that the fitting error does not satisfy the stopping criteria,. And, the apparatus is operable to effect a means for altering the plurality of function parameters and the total number of functions, and operable to repeat the generating, determining, and comparing operations.

20   In a refinement of the foregoing apparatus, it further comprising means for specifying at least a first threshold value delineating the plurality of functions. In another refinement, the at least a first threshold value is calculated based upon the likelihood of misclassification of data. In another refinement, the apparatus further comprises a means for segmenting the data according to the at least a first threshold

25   value. In another refinement, the objective function is defined as a vector representation of the plurality of function parameters. In another refinement, the means for altering operation is accomplished by evolving the plurality of function parameters and the total number of functions according to a genetic algorithm. In another refinement, the genetic algorithm evolves the plurality of function parameters

30   through mutation and crossover. In another refinement, the plurality of functions are

normal distributions, and the plurality of functions parameters include the mean and standard deviations of the normal distributions. In another refinement, the means for comparing includes the utilization of a statistical f-test to evaluate the relative contribution of each of the plurality of functions in comparison of the fitting error and

5   the data. In another refinement, the data is organized as a histogram. In another refinement, the stopping criteria are defined by a fitness function. In another refinement, the fitness function is optimized to minimize the magnitude of the fit error between the objective function and the data.

The present invention also teaches a specific application directed to the

10   field of digital imaging. A method of specifying thresholds for segmenting a digital image is taught. That method comprises the steps of producing a histogram of the image, having histogram data and defining a plurality of functions according to a plurality of function parameters and a total number of functions. Then, generating an objective function based on the plurality of function parameters and determining a

15   fitting error between the objective function and the histogram data. Next, comparing the fitting error to stopping criteria and altering the plurality of function parameters and the total number of functions, and repeating the generating, determining, and comparing steps if the fitting error does not satisfy the stopping criteria. Finally, specifying at least a first threshold value delineating the plurality of functions if the

20   fitting error satisfies the stopping criteria.

In a refinement to the imaging method, the at least a first threshold value is calculated based upon the likelihood of misclassification of the histogram data. In another refinement, the objective function is defined as a vector representation of the plurality of function parameters. In another refinement, the

25   altering step is accomplished by evolving the plurality of function parameters and the total number of functions according to a genetic algorithm. In another refinement, the genetic algorithm evolves the plurality of function parameters through mutation and crossover. In another refinement, the plurality of functions are normal distributions, and the plurality of functions parameters include the mean and standard

30   deviations of the normal distributions. In another refinement, the comparing step

includes the utilization of a statistical f-test to evaluate the relative contribution of each of the plurality of functions in comparison of the fitting error and the data. In another refinement, the stopping criteria are defined by a fitness function. In another refinement, the fitness function is optimized to minimize the magnitude of the fit

5   error between the objective function and the data.

The present invention also teaches an apparatus for determining thresholds for segmenting a digital image. The apparatus includes a means for producing a histogram of the image, having histogram data and a means for defining a plurality of functions according to a plurality of function parameters and a total

10   number of functions. Also, a means for generating an objective function based on the plurality of function parameters and a means for determining a fitting error between the objective function and the histogram data. Also, a means for comparing the fitting error to stopping criteria and a means for altering the plurality of function parameters and the total number of functions, and repeating the generating,

15   determining, and comparing operations if the fitting error does not satisfy the stopping criteria. Finally, a means for specifying at least a first threshold value delineating the plurality of functions if the fitting error satisfies the stopping criteria.

In a refinement to the foregoing apparatus, the at least a first threshold value is calculated based upon the likelihood of misclassification of the histogram

20   data. In another refinement, the objective function is defined as a vector representation of the plurality of function parameters. In another refinement, the altering operation is accomplished by evolving the plurality of function parameters and the total number of functions according to a genetic algorithm. In another refinement, the genetic algorithm evolves the plurality of function parameters through

25   mutation and crossover. In another refinement, the plurality of functions are normal distributions, and the plurality of functions parameters include the mean and standard deviations of the normal distributions. In another refinement, the means for comparing includes the utilization of a statistical f-test to evaluate the relative contribution of each of the plurality of functions in comparison of the fitting error and

30   the data. In another refinement, the stopping criteria are defined by a fitness function.

In another refinement, the fitness function is optimized to minimize the magnitude of the fit error between the objective function and the data.

BRIEF DESCRIPTION OF THE DRAWINGS

5   Figure 1 is a flow diagram of an image segmentation process according to an illustrative embodiment of the present invention.

Figure 2 is an illustration of an image segmentation process according to an illustrative embodiment of the present invention.

Figure 3 is a vector representation of a population mixture model

10   according to an illustrative embodiment of the present invention.

Figure 4 is a basic genetic algorithm flowchart according to an illustrative embodiment of the present invention.

DESCRIPTION OF THE INVENTION

15   Illustrative embodiments and exemplary applications will now be described with reference to the accompanying drawings to disclose the advantageous teachings of the present invention.

While the present invention is described herein with reference to illustrative embodiments for particular applications, it should be understood that the

20   invention is not limited thereto. Those having ordinary skill in the art and access to the teachings provided herein will recognize additional modifications, applications, and embodiments within the scope thereof and additional fields in which the present invention would be of significant utility.

The current invention is a technique for modeling data as the mixture

25   of an indeterminate number of functions. The method employs a global optimization technique, as generally described in the background section of this disclosure, to solve a problem such as the nonlinear least squares problem. This is through utilization of a population mixture model. There are other applications that benefit from population mixture modeling, and these represent other pertinent embodiments of the present

30   invention. By way of example, and not limitation, these include clustering techniques

that attempt to classify points in multiple dimensions as belonging to a given set of clusters. The present invention could readily be used as part of a clustering technique that would not require the number of clusters to be predetermined. Another example (in one-dimension) is found in spectroscopy, where data exists that represents some function of wavelength (usually spectral transmission or reflection). It is desirable to model the data in terms of some mixture of underlying functions (sub-populations). In general, the present invention can be embodied in any application that needs to determine how a given set of data can be described by a mixture of underlying (basis) functions.

As noted herein, segmenting gray scale images is a useful task in imaging science. Image segmentation can be used directly for image compression or as a preprocessing step for background removal, feature detection, pattern recognition algorithms, and so forth. A wide variety of segmentation techniques have emerged in the literature over the last two decades. The simplest methods are point-based, where each pixel is quantized based on thresholds determined by analysis of the histogram of the image.

The process of segmenting an image involves several basic steps, as are illustrated as a process in Figure 1. Such a process starts at step 2 and commences when an image is input at step 4. A histogram is generated at step 6, which is done using traditional techniques as are understood by those of ordinary skill in the art. At step 8, a mathematical model of the histogram data is created using a plurality of probability density functions, which are normal distributions in the preferred embodiment. At step 10, an optimization procedure is used to adjust the mathematical model to more closely fit the histogram data. The optimization is done at a global level. At step 12, a test is made to determine if the optimized model of functions globally fit the actual histogram data to an acceptably close level (otherwise sated as meeting the specified stopping criteria). If not, then the data is again modeled by return to step 8. If the plurality of probability density functions do meet and acceptable level of error (or stopping criteria) respecting the histogram data, the segmentation thresholds are generated at step 14. Finally, the thresholds are applied

to the image data at step 18 and the process is finished at step 18, where the data set may then be used for whatever subsequent process it was intended.

The present invention is directed to the steps of modeling and optimizing the probability density functions, including selection of how many
5   probability density functions (or sub-populations), so as to minimize the fitting error, and yielding an optimized threshold selection.

In its preferred embodiment, the present invention is employed as part of a point-based image segmentation process. Such a process is described by the flowchart in Figure 2. The image 20 to be segmented is analyzed to form a histogram
10 22, depicted by curve 24 (a histogram is typically defined by a discrete sets of values, called 'bins'). The preferred embodiment of the present invention models the histogram data as a mixture of sub-populations, and the model parameters are used to generate thresholds according to some deterministic technique (as described generally in the background section of this disclosure). The thresholds 30 generated in this
15 technique are shown in 26 as overlaying the histogram 28. Based on the determined thresholds, the image is segmented 32 to n levels (three in this example), where n is the number of sub-populations in the population mixture model.

The technique used to globally fit the histogram data includes developing a population mixture model and optimizing it to the data. A population
20 mixture model is a probability density function that can be expressed as the weighted sum of multiple sub-populations (a plurality of other probability density functions). In applications such as image segmentation or cluster analysis, algorithms exist that fit a population mixture model to a given set of data (usually from a histogram). Prior art methods for determining the optimal parameters of the population mixture
25 model require *a priori* knowledge of the number of sub-populations best suited to closely fit the data. This knowledge is frequently attained by manual inspection or by an automatic pre-processing step. Alternatively, optimization techniques utilizing a multi-start approach have been employed in a trial and error fashion. Such methods fit the model repeatedly for different numbers of sub-populations. Multi-start
30 approaches have two major drawbacks. First, the maximum number of sub-

populations is arbitrarily chosen by some preprocessing step, and second, the computational effort required to repeatedly fit the model is immense. The present invention obviates the need for employing a multi-start approach or requiring any *a priori* knowledge by including the number of sub-populations as a parameter in the

5 model and determining the optimal model parameters through the use of a genetic algorithm.

Mathematically, a population mixture model can be modeled as a function of the form:

$$h\left(x, \beta, \alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(n)}\right) = \sum_{i=1}^{n} \beta_i p_i\left(x, \alpha^{(i)}\right) \qquad (1)$$

where each $p_i\left(x, \alpha^{(i)}\right)$ is a probability density function that is described by the parameters in $\alpha^{(i)}$ and defined for every vector x within some feasible region. β is the weighting factor, which varies with each of the p_i's . In the preferred

15 embodiment, the p_i's are taken to be normal distribution probability density functions. In the one-dimensional case, each $\alpha^{(i)}$ would be a two-element vector corresponding to the mean and standard deviation (typically identified as the μ and σ, respectively) of each ith normal distribution. Since the population mixture model is a probability density function, it must integrate to unity. Therefore, the constraint:

20

$$\sum_{i=1}^{n} \beta_i = 1 \qquad (2)$$

must be satisfied.

In order to determine the optimal parameters of the population mixture

25 model for a given set of data (i.e., m ordered pairs $\left(\hat{x}^{(j)}, \hat{h}_j\right)$), a k-dimensional objective function is constructed (where k is the sum of magnitudes of the β and $\alpha^{(i)}$ parameters in Equation (1)), that yields its global optimum, where the population mixture model best fits the data. The global optimum can then be located though a

global optimization procedure. In an illustrative embodiment of the present invention, such an optimization problem is mathematically given by:

$$\text{Minimize} \quad \sum_{j=1}^{m} \left[h\!\left(\hat{x}^{(j)}, \beta, \alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(n)} \right) - \hat{h}_j \right] \tag{3}$$

5

with respect to $\beta, \alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(n)}$

This objective function will be understood by those of ordinary skill in the art to be the well-known sum of squared errors problem, thus classifying Equation (3) as a nonlinear least squares problem. Those of ordinary skill in the art will also

10 generally appreciate the analytical approach to solving least squares problems. Of course, other optimization problems and/or objective functions, such as maximum likelihood functions, can be used, and fall within the scope of the present invention.

In point-based segmentation methods, the value $I(u,v)$ at each pixel are assumed to be drawn from one of n sub-populations. As noted above, these sub-

15 populations are normal distributions in the preferred embodiment. The histogram of the image will then appear to be a mixture of the sub-populations, as modeled in Equation (1).

Given that Equation (3) has been solved (which process will be more fully developed herein after), including a determination of the number of probability

20 density functions (n), an optimal set of thresholds T can be calculation by minimizing classification error. The $n = 2$ case is derived in Gonzalez, R. C. and Wintz, P. *Digital Image Processing.* Addison-Wesley, 1977.

Gonzalez and Wintz teach that if it is known *a priori* that an image contains only two principal brightness regions, then the histogram of that image is be

25 considered as an estimate of the brightness probability density function. The overall density function would be the sum or mixture of two unimodal densities, one for the bright region and one for the dark region in the image. Furthermore, the mixture parameters would be proportional to the areas of the picture for each brightness. If the form of the densities is known or assumed, then it is possible to determine an

optimum threshold (in terms of minimum error) for segmenting the image into the two brightness regions.

Further, suppose that an image contains two values combined with additive Gaussian noise. The mixture probability density function is given by:

$$p(x) = P_1 p_1(x) + P_2 p_2(x) \tag{4}$$

which, for the Gaussian case, is:

$$p(x) = \frac{P_1}{\sqrt{2\pi}\sigma_1} \exp\frac{(x - \mu_1)^2}{2\sigma_1^2} + \frac{P_2}{\sqrt{2\pi}\sigma_2} \exp\frac{(x - \mu_2)^2}{2\sigma_2^2} \tag{5}$$

where μ_1 and μ_2 are the mean values of the two brightness levels, σ_1 and σ_2 are the standard deviations about the means, and P_1 and P_2 are the *a priori* probabilities of the two levels. Since the constraint:

$$P_1 + P_2 = 1 \tag{6}$$

must be satisfied, the mixture density has five unknown parameters. If all the parameters are known, the optimum threshold is easily determined.

Assuming that the dark region corresponds to the background and the bright region corresponds to objects in the image, then $\mu_1 < \mu_2$ and a threshold, T, may be defined so that all pixels with brightness level below T are considered background points and all pixels with level above T are considered object points. The probability of erroneously classifying an object point as a background point is:

$$E_1(T) = \int_{-\infty}^{T} p_2(x)dx \tag{7}$$

Similarly, the probability of classifying a background point as an object point is:

$$E_2(T) = \int_T^\infty p_1(x)dx \qquad (8)$$

Therefore, the overall probability of error is given by:

$$E(T) = P_2 E_1(T) + P_1 E_2(T) \qquad (9)$$

To find the threshold value for which this error is minimum, differentiate $E(T)$ with respect to T (using Liebnitz's rule) and equate the result to zero. The result is then:

$$P_1 p_1(T) = P_2 p_2(T) \qquad (10)$$

Applying this result to the Gaussian density gives, after taking logarithms and simplifying, a quadratic equation:

$$AT^2 + BT + C = 0 \qquad (11)$$

where:

$$A = \sigma_1^2 - \sigma_2^2 \qquad (12)$$

$$B = 2\left(\mu_1 \sigma_2^2 - \mu_2 \sigma_1^2\right) \qquad (13)$$

$$C = \sigma_1^2 \mu_2^2 - \sigma_2^2 \mu_1^2 + \sigma_1^2 \mu_2^2 \ln\left(\sigma_1 P_1 / \sigma_2 P_2\right) \qquad (14)$$

The possibility of two solutions indicates that it may require two threshold values to obtain the optimum solution. If the variances are equal, $\sigma^2 = \sigma_1^2 = \sigma_2^2$, a single threshold is sufficient:

$$T = \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2}{\mu_1 - \mu_2}\ln\left(\frac{P_2}{P_1}\right) \qquad (15)$$

If the prior probabilities are equal, $P_1 = P_2$, the optimum threshold is just the average of the means. The determination of the optimum threshold may be accomplished easily for other unimodal densities of known form, such as the Raleigh and log-normal densities.

To estimate the parameters from a histogram of an image one may use a maximum likelihood or minimum mean-square error approach. For example, the mean-square error between the mixture density, $p(x)$, and the experimental histogram $h(x)$, is:

$$M = \frac{1}{N}\sum_{i=1}^{N}\{p(x_i) - h(x_i)\}^2 \qquad (16)$$

where an N-point histogram is available.

Gonzalez and Wintz also note that in general, it is not a simple matter to determine analytically parameters that minimize this mean-square error. Even for the Gaussian case, the straightforward computation of equating the partial derivatives to zero leads to a set of simultaneous transcendental equations that usually can be solved only by numerical procedures. Since the gradient is easily computed, a conjugate gradient, or Newton's method for simultaneous nonlinear equations, may be used to minimize M. With either of these iterative methods, starting values must be specified. Assuming the *a priori* probabilities to be equal may be sufficient. Starting values for the means and variances may be determined by detecting modes in the histogram or simply dividing the histogram into two parts about its mean value, and computing means and variances of the two parts to be used as starting values.

The foregoing example teaches the n=2 case for threshold derivation, this is but one of many possibilities numbers of probability density functions that may

be applicable. As is understood by those of ordinary skill in the art, the general case is an extension of the $n=2$ case.

The general case of image segmentation thresholding will now be further developed. Define $\mathbf{T}$, the set of thresholds, $\mathbf{T} = (T_1, T_2, \ldots, T_{n-1})^{\mathbf{T}}$ and assume

5   that $T_0 < T_1 < T_2 < \ldots < T_{n-1} < T_n$, where $T_0 = \min_j \hat{x}_j$ and $T_n = \max_j \hat{x}_j$. Given a set of thresholds, an image I is segmented to form a new image $\hat{I}$ as follows:

$$\hat{I}(u,v) = \begin{cases} \phi(1), & T_0 \le I(u,v) < T_1 \\ \phi(2), & T_1 \le I(u,v) < T_2 \\ \vdots & \vdots \\ \phi(n) & T_{n-1} \le I(u,v) \le T_n \end{cases}, \quad (17)$$

10   where ϕ is an arbitrary one-to-one function, such as $\phi(i) = i$ or $\phi(i) = \mu_i$.

To determine the optimum set of thresholds given β^*, μ^*, and σ^*, first introduce two random variables. Let $\mathbf{X}$ be the index of the sub-population to which a given pixel belongs, and let $\mathbf{C}$ be the index of the segment into which it is

15   classified. Then the optimal $\mathbf{T}$ vector will maximize the probability that a pixel is correctly classified (or minimize the probability that a pixel is misclassified). Then:

$$\theta(\mathbf{T}) = P\{\text{misclassification}\} = 1 - P\{\text{correct classification}\}$$
$$= 1 - \sum_{i=1}^{n} P\{\mathbf{C} = i, \mathbf{X} = i\} \quad (18)$$
$$= 1 - \sum_{i=1}^{n} P\{\mathbf{C} = i \mid \mathbf{X} = i\} P\{\mathbf{X} = i\}$$

20   It is know that $P\{\mathbf{X} = i\} = \beta_i^*$ and $P\{\mathbf{C} = i \mid \mathbf{X} = i\} = \int_{T_{i-1}}^{T_i} p_i(x)dx$, therefore:

$$\theta(\mathbf{T}) = 1 - \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \beta_i^* p_i(x)dx. \quad (19)$$

Any point that minimizes Equation (19) satisfies $\nabla\theta(\mathbf{T}) = 0$. From the Fundamental Theorem of Calculus, then:

$$\nabla\theta(\mathbf{T}) = \begin{pmatrix} \beta_2^* p_2(T_1) - \beta_1^* p_1(T_1) \\ \beta_3^* p_3(T_2) - \beta_2^* p_2(T_2) \\ \vdots \\ \beta_n^* p_n(T_{n-1}) - \beta_{n-1}^* p_{n-1}(T_{n-1}) \end{pmatrix} \tag{20}$$

5

Therefore, any optimal vector $\mathbf{T}^*$ satisfies $\beta_i^* p_i(T_i^*) = \beta_{i+1}^* p_{i+1}(T_i^*)$ for $i = 1, ..., n-1$. Taking logarithms of both sides of these equations and simplifying yields quadratic equations in $\mathbf{T}^*$:

$$\left(\sigma_i^{*2} - \sigma_{i+1}^{*2}\right)T_i^{*2} + 2\left(\mu_i^* \sigma_{i+1}^{*2} - \mu_{i+1}^* \sigma_i^{*2}\right)T_i^* + \mu_{i+1}^{*2}\sigma_i^{*2}$$
$$- \mu_i^{*2}\sigma_{i+1}^{*2} + \sigma_i^{*2}\sigma_{i+1}^{*2}\ln\left(\frac{\sigma_i^* \beta_i^*}{\sigma_{i+1}^* \beta_{i+1}^*}\right) = 0 \tag{21}$$

10

Therefore, to determine a set of optimal thresholds given the population mixture model, solve the quadratic equations given by Equation (21). It can be shown that any solution to Equation (21) that satisfies

15 $\mu_1^* < T_1^* < \mu_2^* < T_2^* < \cdots < T_{n-1}^* < \mu_n^*$ also satisfies the second-order conditions ($\nabla^2\theta(\mathbf{T})$ positive definite), and hence is a local minimizer of Equation (19). The global minimizer can be found by finding the best of all the local minima.

Since it has now been demonstrated that the determination of thresholds is a straightforward deterministic process, the difficulty in the prior art

20 with point-based methods is the determination of a population mixture model, or, otherwise stated, the difficulty is in finding a solution to Equation (3). As noted earlier, many local minima typically exist for this kind problem, so attacking it with simple local optimizers will not be adequate. Also, since the problem can potentially have high dimensionality, none of the traditional deterministic methods appear

25 promising, except for the trivial cases. However, as mentioned above, some of the Monte Carlo techniques can be used to solve Equation (3). In an illustrative

embodiment of the present invention, a genetic algorithm is utilized to solve Equation

(3), and particularly in the case when n is not chosen, or otherwise known, *a priori*.

For fixed n, available *a priori*, any of a variety of global optimization

techniques can be used to solve Equation (3), see generally Global Optimization:

5   Techniques and Applications N. Cahill, RIT Master's Thesis, May 2000, the contents

of which are hereby incorporated by reference thereto. In the context of image

segmentation, Snyder, Bilbro, Logenthiran, and Rajala (Optimal Thresholding - A

New Approach, Pattern Recognition Letters, 11:803-810, 1990) describe a tree

annealing technique that solves Equation (3). However, their method requires an

10   immense amount of storage in computer memory and can be extremely inefficient. In

a practical sense, this is unacceptable as the utilization of such techniques is naturally

constrained by the processing requirements each bears. One way to make such an

approach more efficient is to recognize that the population mixture model Equation

(1) is linear with respect to the β parameters.

15   Since Equation (1) is linear with respect to the β parameters, then

given any set of $\alpha^{(i)}$'s, the estimation of the β parameters can be separated from the

global optimization procedure, and can be computed using a deterministic method

(Equation (3) becomes a least squares with equality constraints problem). Those of

ordinary skill in the art will appreciate that this effectively reduces the k-dimensional

20   objective function to a $(k-n)$-dimensional objective function. Thus, implying a

significant reduction in processor overhead for real-world implementations.

The problem with the techniques described in the previous paragraphs

is that the number of sub-populations n must be fixed and known *a priori*. This

requires either some preprocessing step to automatically determine n, or some multi-

25   start optimization technique that solves Equation (3) for a sequence of different

values for n. Preprocessing techniques tend to be *ad hoc*, and, multistart procedures

are extremely time, and processor resource, consuming. The present invention

combines determination of the number of sub-populations with the objective function

evaluation process and estimates n as part of the global optimization procedure.

The global optimization procedure used in the preferred embodiment of the present invention utilizes a genetic algorithm. Genetic algorithms are described by Holland in Adaptation in Natural and Artificial Systems, The University of Michigan Press, Ann Arbor, 1975, the teachings of which are hereby incorporated

5   herein by reference. One specific application of genetic algorithms is described in US patent 6,004,015 to Watenabe for OPTIMIZATION ADJUSTING METHOD AND OPTIMIZATION ADJUSTING APPARATUS, the teachings of which are hereby incorporated herein by reference. Another application of genetic algorithms used to solve a large class of problems is detailed in US patent 5,136,686 to Koza for NON-

10   LINEAR GENETIC ALGORITHMS FOR SOLVING PROBLEMS BY FITTING A FIT COMPOSITION OF FUNCTIONS, the teachings of which are hereby incorporated herein by reference.

Consideration is now directed to the process of estimating population mixtures. That is, the nonlinear least squares problem, Equation (3), which estimates

15   the population mixture parameters for a given image histogram. For reference, Equation (3) appears again:

$$\text{Minimize} \quad \sum_{j=1}^{m} \left[h\left(\hat{x}^{(j)}, \beta, \alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(n)} \right) - \hat{h}_j \right] \qquad (3)$$

with respect to $\beta, \alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(n)}$

20   For fixed μ and σ, which are the mean and standard deviation for the normal distributions, Equation (3) is a linear least squares problem in terms of β. Taking into account the equality constraint given in Equation (2) and the inequality constraint $\beta \geq 0$, this is a standard least squares equality constraint problem (hereinafter 'LSEI') that can be solved readily for β. Therefore, Equation (3) can be

25   solved by partitioned least squares, using Gauss-Newton to update estimates for the nonlinear parameters μ and σ and solving the resulting LSEI problem for β at each iteration. This effectively reduces the dimensionality of the nonlinear least squares problem by one third.

It is extremely difficult to find the global minimum of Equation (3), even if the problem is partitioned. If there are n modes in the population, Equation (3) describes a function in $3n$-dimensional space ($2n$ for the partitioned problem) that has a large number of local minima. Because of the relatively high dimensionality of this problem, deterministic global optimization techniques are impracticable. Therefore, Monte Carlo techniques are the only reasonable alternative. A multistart method could theoretically work, however, this is inefficient because nothing is known about the basins of attraction of the various local minima. Another possibility is simulated annealing, and this problem appears to lend itself to such an approach.

A solution to Equation (3) is described by Snyder, W., Bilbro, G., Logenthiran, A., and Rajala, S. in Optimal Thresholding - A New Approach. *Pattern Recognition Letters*, 11:803-810, 1990, the contents of which are hereby incorporated by reference thereto. That solution is via tree annealing, which is a variant of simulated annealing. In tree annealing, the search space is described by a tree, where each level of the tree describes some binary partition of one of the parameters. For Equation (3), the search space is described by $0 \leq \mu \leq 1$, $\alpha_1 \leq \sigma \leq \alpha_2$, and $0 \leq \beta \leq 1$, where α_1 and α_2 are suitable lower and upper bounds on the standard deviations. The annealing procedure takes place by randomly traveling down one branch of the tree and then comparing the "leaf" point with the current iterate according to the Metropolis criterion. If the leaf point is chosen as the next iterate, another node is placed at that leaf to increase the resolution of the tree. If the Metropolis criterion is changed according to some cooling schedule, over time the tree will be highly resolved in areas of local minima.

There are two major drawbacks to the tree annealing approach. First, the size of the tree becomes enormous and storage becomes impractical, and second, it is not readily apparent how to satisfy the equality constraint in Equation (2). The first drawback can be slightly alleviated by solving the partitioned problem. Even though the dimensionality of the problem is somewhat reduced, it does not become small enough to disregard the issue of tree size. The second drawback can be ignored

if a standard elimination or transformation method is used to remove the equality constraint. However, the transformed parameters will have no intuitive significance.

Another technique for histogram thresholding is given by Brunelli, R. Optimal Histogram Partitioning Using a Simulated Annealing Technique, *Pattern Recognition Letters*, 13:581-586, 1992, the contents of which are hereby incorporated by reference thereto. Simulated annealing does not attempt to fit a population model to the histogram, rather, it solves a combinatorial optimization problem via simulated annealing that yields the segmentation thresholds directly. The benefit of using this method is that it is not required (although it is useful) to determine the number of modes/thresholds prior to performing the optimization. On the other hand, this method does not yield any useful information for estimating sub-population parameters.

In Brunelli's method, the image histogram is preprocessed to locate every local minimum. A node is defined by the n local minima that are found. Seven pieces of information are stored at each node: node location, histogram value, and cumulative distribution value, location, histogram value, and cumulative distribution value of the nearest local maximum to the right of the node, and a binary flag. A configuration x is defined as the n-vector of all binary flags. Every element of x that is a one is considered to be a threshold value. There are 2^n possible configurations, one of which describes the optimal set of thresholds. The objective function to be minimize is then given by:

$$f(\mathbf{x}) = -\sum_s \frac{\min\left(|m_s - k_{s-}|, |m_s - k_{s+}|\right)}{m_s}\left(1 - \frac{A_s}{m_s d_s}\right),$$
(22)

where s is a slice of the current configuration (a segment of the histogram corresponding to the interval containing a node $x_i = 1$ and the nearest node to its right with $x_j = 1$), m_s is the maximum histogram value of the slice, k_{s-} is the histogram value of the left node, k_{s+} is the histogram value of the right node, A_s is the area of the slice, and d_s is the width of the slice. Since the nearest maximum and the

cumulative distribution values are stored at each node, computing m_S and A_S is simple and efficient.

If the number of modes of the histogram is known *a priori* (say *m* modes), Equation (22) can be transformed to favor configurations with *m* active by

5   defining a new objective function:

$$\hat{f}(x) = \frac{f(x)}{\alpha \left| \sum_{i=1}^{n} x_i - m \right| + 1}, \qquad (23)$$

where α is a weighting factor.

10   Simulated annealing can now be used to minimize Equation (22) or Equation (23). New configurations in a "neighborhood" of the current configuration can be generated by flipping the active status of a uniformly distributed (from 1 to *n*) random number of bits. Brunelli teaches the use of both the geometric cooling schedule and an exponentially decaying cooling schedule, performing the Metropolis

15   algorithm ten thousand times at each temperature level.

Both of the previously described methods for histogram thresholding have their strengths and weaknesses. Snyder's method (extended to solve the partitioned least squares problem) does a good job at finding the optimal configuration of parameters for the model in Equation (1), but it requires *a priori*

20   knowledge of the number of sub-populations. Brunelli's method performs reasonably well when the number of sub-populations or desired thresholds is not known, but the notion of "goodness of shape" that it relies on is sketchy. The present invention image segmentation strategy uses genetic algorithms to find the optimal set of parameters for the histogram model in Equation (1) without requiring that the number

25   of sub-populations *n* be predetermined.

The present invention teaches a genetic evolution technique for histogram thresholding, including descriptions of the encoding of chromosomes, the crossover, mutation, and selection operators, and the fitness function. Any global optimization technique that is used to find the optimal parameters of the population

mixture model requires the formal mathematical definition of a specific collection of parameters. In the preferred global optimization technique, this collection of parameters is called a chromosome. Figure 3 illustrates a specific chromosome x 40 that describes a collection of sub-populations 42, 44, 46, which are probability

5 density functions 34, 36, 38 in the preferred embodiment. The values are shown as variables A 42, B 44, and C 46 in Figure 3. Normal distributions were selected for the probability density functions. Each normal distribution is parameterized by its mean (μ) and standard deviation (σ).

Each chromosome is a length b vector 40 that defines a collection of

10 Gaussians (normal distributions), where b is equal to the number of histogram bins. For a given chromosome x, if $x_i \neq 0$, then that configuration contains a Gaussian with mean located at the ith bin. (If $x_i = 0$, then that configuration contains no Gaussian centered at the ith bin). The value of x_i is the standard deviation of that Gaussian, identified by A, B, and C in Figure 3. Therefore, if an appropriate range for standard

15 deviation values is defined (say $0.01 \leq \sigma \leq 0.5$) and specify a vector of c linearly spaced values spanning this range, then there exist $b(c+1)$ possible chromosomes that provide good coverage of the μ and σ spaces for any $n \leq b$. β can be found by solving the LSEI problem arising from Equation (3), so β is implicitly stored for each chromosome.

20 With a rule that defines each specific mixture of sub-populations as a chromosome, a genetic algorithm can be utilized to find the chromosome that has the best fitness (i.e. is a global optimum of the objective function). The basic flow of a genetic algorithm is shown in Figure 4. The process begins at step 48 and proceeds to step 50 where an initial generation (collection of chromosomes) is chosen. At step

25 52, the fitness values (objective function values) are computed for each chromosome. The stopping criteria are checked at step 54, which entails a determination if the fitting error is acceptably low. If it is, then the process is exited at step 62. Otherwise, if the stopping criteria are not met at step 54, then individual chromosomes are selected for reproduction at step 56. Reproduction is simulated by

30 crossover and mutation operators at step 58. A final population is thus formed at step

60, and, and the process is repeated by replacing the initial generation of chromosomes from step 50 with the final population from step 60 and repeating the process from step 52. Many varieties of the basic genetic algorithm have arisen in the literature as noted herein before. But, it will be understood by those skilled in the art

5   that the use of a different optimization technique to find the optimal parameters of the population mixture model, with an indefinite number of parameters, will not deviate from the spirit or scope of this invention.

In the preferred embodiment, single-point genetic crossover (with probability of crossover $p_c = 0.7$) is suggested as sufficient. After two chromosomes

10   are crossed over, the mutation operator is applied to the offspring. Each allele (an allele is a single element of chromosome x) is mutated with probability $p_m = 0.001$. Since the chromosomes are not binary vectors, the mutation of an allele is defined by the function $\omega(x)$:

15

$$\omega(x) = \begin{cases} \kappa, & x = 0 \\ 0, & x \neq 0 \end{cases}, \qquad (24)$$

where κ is a random variable uniformly chosen from the possible values of σ. Equation (24) allows a new sub-population to be entered into the mixture or a current sub-population to be removed, thereby effectively changing the number of probability

20   density functions, n. Given a current generation of chromosomes, selection occurs by probabilities weighted by the inverse of the relative fitness of the chromosomes. The inverse is taken since the goal is to minimize, not maximize, the objective function. An initial population of N chromosomes ($N = 25$, 50, or 100 are reasonable choices in the preferred embodiment) must be chosen so that it provides adequate coverage of

25   parameter space. Since it is assumed that each histogram has at least two sub-populations (otherwise segmentation would be meaningless), an initial chromosome x is generated by randomly choosing the number of modes n and then randomly choosing n elements of x in which to place nonzero values. The nonzero values are chosen uniformly from the set of possible standard deviations. The random choice of

n can be achieved by assigning $n = 2 + \rho$, where $\rho \sim$ Poisson(λ). In the preferred embodiment, a reasonable choice for the mean of ρ is $\lambda = 3$.

To fully describe this technique for histogram thresholding, all that remains is the definition of the fitness function. As noted above, for this problem, a

5 fitness function is defined that not only takes residual error into account, but also the number of sub-populations required to adequately explain variation. Through the use of the crossover and mutation operators, the genetic algorithm can mix and match sub-populations, effectively treating the number of sub-populations in the population mixture model as another parameter. This process does require modification to the

10 objective function (Equation 3). If the objective function only computes some error measure, such as the sum of squared errors between the data and the model as shown in Equation 3, the minimum error will naturally occur when the maximum number of sub-populations are used. Therefore, in the preferred embodiment, the objective function must include both some error measure and some notion of the relative

15 contributions of the individual sub-populations to the model fit. The preferred way of including these contributions is to use a statistical F-test to reduce the error term in the objective function when the number of sub-populations in the mixture model is adequate. This is accomplished by computing the error in a "master chromosome" model. This means computing the error by the model containing all sub-populations

20 in a given generation of the genetic algorithm. In order to do this, divide the fitness computation into two parts. When a given chromosome is created, calculate the residuals of the corresponding population model and determine the sum squared error (sse) as follows:

25
$$sse(\mathbf{x}) = \sum_{j=1}^{m} \left[h(\hat{x}_j, \beta_x, \mu_x, \sigma_x) - \hat{h}_j \right] \tag{25}$$

where β_x, μ_x, and σ_x are the parameters of the histogram model Equation (1) associated with the chromosome $\mathbf{x}$, and the m ordered pairs $(\hat{x}_j, \hat{h}_j)$ compose the histogram. This calculation is performed for each chromosome in a given generation.

Once the entire generation has been created, a master chromosome **M** that contains every unique (μ, σ) pair that exists in any chromosome of the generation is formed. Note that it may not be possible to explicitly code **M** as a chromosome because there may be multiple Gaussians with the same mean. It is not necessary to do this,

5 however, all that is needed is to store $\mu_{\mathbf{M}}$ and $\sigma_{\mathbf{M}}$, the means and standard deviations of all Gaussians described by **M**. Given $\mu_{\mathbf{M}}$ and $\sigma_{\mathbf{M}}$, LSEI problem can be solved (Equation (3)) for $\beta_{\mathbf{M}}$. In the statistical sense, $h(x, \beta_{\mathbf{M}}, \mu_{\mathbf{M}}, \sigma_{\mathbf{M}})$ can be thought of as the "full" model and $h(x, \beta_x, \mu_x, \sigma_x)$ as a reduced model of the histogram. Using an F test, a fitness function $f(\mathbf{x})$ can be defined that accounts for both residual error and

10 for the ability to explain variation:

$$f(\mathbf{x}) = \frac{sse(\mathbf{x})}{F_{\alpha, k-n, m-(k+1)}} \left[\frac{(sse(\mathbf{x}) - sse(\mathbf{M}))(m - (k+1))}{sse(\mathbf{M})(k - n)} \right] \tag{26}$$

where k is the number of modes in **M** and $F_{\alpha, k-n, m-(k+1)}$ is the critical value for an F

15 distribution with parameters $\alpha (= 0.05)$, $k\text{-}n$, and $m\text{-}(k{+}1)$. The effect of Equation (26) is to increase Equation (25) when the hypothesis that the reduced model is acceptable would be rejected and to decrease Equation (25) when this hypothesis would not be rejected. Equation (26) is driven to a minimum when both the residuals are small and n is not deemed unacceptable.

20 However, two problems can potentially arise. First, the number of modes in **M** can exceed the number of data points m extracted from the histogram. If this occurs in a given generation, the histogram is resampled so that as many points are needed are extracted. The second problem is that if two or more modes in **M** are similar (means and standard deviations are close), the predictor matrix in the LSEI

25 problem can be poorly scaled. If this is the case, it is acceptable to remove modes from **M** that are nearly linearly dependent on other modes.

Thus, the present invention has been described herein with reference to a particular embodiment for a particular application. Those having ordinary skill in

the art and access to the present teachings will recognize additional modifications, applications and embodiments within the scope thereof.

It is therefore intended by the appended claims to cover any and all such applications, modifications and embodiments within the scope of the present invention.